MEG Decoding of Happy Faces through Spatial Covariance Matrices

Emanuele Olivetti and Paolo Avesani

NeuroInformatics Laboratory (NILab), Bruno Kessler Foundation, Trento, Italy

Center for Mind and Brain Sciences (CIMeC), University of Trento, Italy

Abstract

This document describes the technical details of our submission to the Data Analysis Competition 3 of Biomag 2016, which requires to predict, from MEG data, in which trials the subjects were presented happy faces. The method proposed in this submission comprises three steps: 1) representing each trial as the covariance matrix of the signals measured by 204 MEG gradiometers, 2) representing each covariance matrix in the tangent space of the associated Riemannian manifold, 3) training a logistic regression classifier on such vectorial representations in order to predict the probability of happy faces on the test set. Over the training data from the 4 subjects provided by the Competition, the proposed method showed very accurate decoding, by scoring an average cross-validated ROC AUC of 0.93.

1 Introduction

Brief summary of the Competition The Data Analysis Competition 3 of Biomag2016 provides MEG data from 4 subjects recorded while presenting visual stimuli, i.e. faces with 6 different facial expressions. The subjects had to press a button when presented a happy face. The Competition provides the class labels of the stimuli for the first half of the recording session, for each subject. The task of the Competition is to predict in which trials of the second half of the session the subject was presented a happy face, just from MEG data.

Brief description of our submission The method described in this submission is based on three main steps:

1. Representing each trial as the spatial covariance matrix of the measured signals, as in [1, 2].

2. Representing each covariance matrix in the tangent space of the associated Riemannian manifold, in order to obtain an effective vectorial representation.

3. Training a logistic regression classifier on such vectorial representation of the train set, in order to predict the probability of happy faces on the test set.

Preview of results and main message The proposed method, applied to the training data of Competition 3, provided accurate decoding of the target stimulus, i.e. happy face vs. other facial expressions, for all the 4 subjects. The cross-validated ROC AUC score, averaged over the subjects, was 0.93.

summary of the report In the following, in Section 2, we introduce the main ingredients of the proposed method. Then, in Section 3, we provide all the details of our analysis, in order to replicate our results, together with the results. In Section 4, we discuss the results and mention future work.

2 Methods

In this section we describe some theoretical ingredients of the proposed method, namely the representation of trials as spatial covariance matrices and some tools of Riemannian geometry to manipulate such matrices.

2.1 Spatial Covariance Matrix (SCM)

Let $\mathbf{x}_t \in \mathbf{R}^n$ the MEG signal recorded at time t from n channels. The spatial covariance matrix (SCM) is defined as

$$\mathbf{\Sigma} = E[(\mathbf{x}_t - E[\mathbf{x}_t])(\mathbf{x}_t - E[\mathbf{x}_t])^\top]. \tag{1}$$

The SCM is a symmetric and positive definite (SPD) matrix. The SCM of each trial can be estimated from the recorded MEG data. Notice that MEG data of single trials may be rank deficient, so $\mathbf{\Sigma}$ may need to be estimated with some form of regularization. In our analysis, we estimated the covariance matrix of each trial using the Oracle Approximating Shrinkage algorithm proposed in [4].

2.2 SCMs on the Riemannian manifold: Tangent Space

In order to use SCMs in the decoding framework, i.e. with classification algorithms, we need a proper notion of distance between SCMs or an accurate vectorial representation, depending on the specific classification algorithm. Typically, in classification problems where matrices are used to represent examples, the Frobenius norm is used as matrix distance and the unfolding of the matrix into a vector, called vectorization ($vec(\cdot)$), is the standard vector representation.

As explained in [1, 2], the fact that SCMs are SPD requires a different notion of distance and vector representation, because the standard Frobenius norm and the simple matrix vectorization do not properly handle the SPD constraint, leading to inaccurate or incorrect results. The space of SPDs of size $n \times n$ is a differentiable manifold $\mathcal{M}$ of dimension $\frac{n(n+1)}{2}$ and the manipulation of such matrices requires the tools of Riemannian geometry [3]. Here we briefly mention the two main results useful for our analysis and refer the interested readers to the references mentioned before for any additional information.

2.2.1 Distance between SCMs

The Riemannian distance between two SPD matrices $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ is

$$\delta(\mathbf{\Sigma}_1, \mathbf{\Sigma}_2) = \| \log(\mathbf{\Sigma}_1^{-\frac{1}{2}} \mathbf{\Sigma}_2 \mathbf{\Sigma}_1^{-\frac{1}{2}}) \|_F \tag{2}$$

where

$$\mathbf{\Sigma}^\alpha = \mathbf{U} \operatorname{diag}(\sigma_1^\alpha, \ldots, \sigma_n^\alpha) \mathbf{U}^\top \tag{3}$$

$$\log \mathbf{\Sigma} = \mathbf{U} \operatorname{diag}(\log(\sigma_1), \ldots \log(\sigma_n)) \mathbf{U}^\top \tag{4}$$

$\mathbf{\Sigma} = \mathbf{U} \operatorname{diag}(\sigma_1, \ldots \sigma_n) \mathbf{U}^\top$ is the eigenvalue decomposition and $\| \cdot \|_F$ is the Frobenius norm.

2.2.2 Vectorial representation of SCMs

For each SCM on the Riemannian manifold, it is possible to define a vectorial tangent space to that point. By projecting other SCMs on that space, we obtain an approximate vectorial representation of the covariance matrices, which is accurate if the SCMs are located in a small neighborhood of the initial SCM. For this reason, given a set of SCMs, their mean SCM is usually used an center point for the tangent space. The mean of a set of m SCMs is defined similarly to the standard geometric mean:

$$\mathbf{\Sigma}_\mu = \operatorname{mean}(\mathbf{\Sigma}_1, \ldots, \mathbf{\Sigma}_m) = \arg\min_{\mathbf{\Sigma} \in \mathcal{M}} \sum_{i=1}^{m} \delta^2(\mathbf{\Sigma}, \mathbf{\Sigma}_i) \tag{5}$$

which is unique for the space of SPD matrices $\mathcal{M}$ and can be estimated with an iterative algorithm (see [1]). Once the $\mathbf{\Sigma}_\mu$ is estimated, the associated vectorial representation $\mathbf{s}$ of an SCM $\mathbf{\Sigma}$, i.e. its projection on the vectorial space centered in $\mathbf{\Sigma}_\mu$, is

$$\mathbf{s} = \operatorname{Vec}(\log(\mathbf{\Sigma}_\mu^{-\frac{1}{2}} \mathbf{\Sigma} \mathbf{\Sigma}_\mu^{-\frac{1}{2}})). \tag{6}$$

where $\operatorname{Vec}(\mathbf{\Sigma})$ is vector of $\frac{n(n+1)}{2}$ dimensions obtained by unfolding of the upper triangular part of the symmetric matrix $\mathbf{\Sigma}$, multiplying by $\sqrt{2}$ and then by concatenating the diagonal elements. In this way and under the neighborhood assumption, $\delta(\mathbf{\Sigma}_A, \mathbf{\Sigma}_B) \approx \|\mathbf{s}_A - \mathbf{s}_B\|_2$.

3 Analysis and Results

In this section we briefly describe the dataset of Competition 3 of Biomag2016, provide the details of our analysis through the proposed method and report the results. All the code or our analysis was developed in Python language[1], on top of the libraries NumPy[2], Scikit-learn[3] and pyRiemann[4] and it is available under a Free/OpenSource license at `https://github.com/emanuele/biomag2016_Competition3`.

3.1 Biomag2016 Competition 3: Dataset

The Competition dataset comes from an experiment using MEG recordings[5] where subjects did pay attention to a stream of images: presentation rate=$1Hz$, stimulus onset asynchrony=$1000ms$, stimulus duration=$333ms$ each block of 12 images contains faces from the same person, but with different facial expressions corresponding to 6 different classes: anger, disgust, fear, neutrality, sadness, and happiness. The data from four healthy adult subjects (mean age=33.8, 3 males) is pre-processed using bandpass filtering and maxfilter. The original data was acquired at $1kHz$ with a 306 channel MEG system (Elekta Neuromag), which is comprised of 204 planar gradiometers and 102 magnetometers. The signal was bandpassed between 0.1 and $41Hz$, and downsampled to $125Hz$.

3.2 Details of the Method and Results.

Extraction of Trials Our analysis started by extracting the relevant time window for each trial from the continuous recording of the Competition dataset. Such window started at $120ms$ (15 timesteps) from the stimulus onset and lasted until $1320ms$ (165 timesteps) from stimulus onset, i.e. $T = 1200ms$ (150 timesteps). Such values were defined in order to take into account high-level visual processing and premotor/motor activation during the experiment. Moreover they gave the highest cross-validated ROC AUC on the train set of all subjects (see later). After trial extraction, the Competition dataset became a 3D matrix of real numbers of size $480 \times 204 \times 150$ (trials $\times$ channels $\times$ timesteps), for each subject. For the first 240 trials, i.e. the first part of the extracted matrix ($240 \times 204 \times 150$), binary class labels, namely happy faces vs. other facial expressions, were provided by the Competition[6]. Those trials and their labels corresponded to the train set. The second part of the extracted matrix ($240 \times 204 \times 150$) corresponded to the test set, for which such class labels had to be predicted.

[1] `https://www.python.org`

[2] `http://www.numpy.org`

[3] `http://scikit-learn.org`

[4] `https://github.com/alexandrebarachant/pyRiemann`

[5] The description is taken from `https://sites.google.com/site/hubertcecotti/home/biomag2016`.

[6] The binary class labels are imbalanced, because a happy face occurs once every six stimuli, on average.

	Subj 1	Subj 2	Subj3	Subj4	Average
ROC AUC	1.0	0.85	0.92	0.98	**0.93**

Table 1: 10-fold cross-validated ROC AUC over the training set of each subject. The average across subjects is reported in bold face.

SCM: estimation For each trial, we estimated the SCM using the Oracle Approximating Shrinkage estimator described in [4] and available from Scikit-learn/pyRiemann. Such estimator provides regularization which is needed because MEG data may be rank deficient, also because of the use of maxfilter in the preprocessing.

SCM: vectorial representation For each subject, the whole set of trials, i.e. train set and test set, was used in order to estimate the mean SCM, to construct the tangent space and to obtain the vectorial representation of all trials, as explained in Section 2.2.2. The Python package pyRiemann provided the necessary tools. For each subject, the resulting dataset became a matrix 480×20910 (trials $\times$ vectors of size $\frac{n(n+1)}{2}$, $n = 204$).

Logistic Regression: cross-validation and testset prediction With trials now represented as vectors, first we estimated the 10-fold cross-validated ROC AUC on the train set[7], using the logistic regression algorithm, by predicting the probability of each trial of being a happy face. The regularization parameter of logistic regression was estimated by cross-validation to be around 10^4. Then, we trained a new logistic regression classifier on the whole train set and predicted the test set with it, in order to create our submission to the Competition.

Results cross-validated on the train set The results of our pipeline of analysis are summarized in Table 1, where the 10-fold cross-validated ROC AUC over the training set of each subject is reported, resulting in an average of 0.93. For each subject, the overall execution time of the code lasted several minutes, on a modern laptop. The main part of the computation was the creation of the mean SCM, which is an optimization problem based on gradient descent. The other parts of the computation required just a few seconds.

4 Discussion and Conclusions

main comment to the results In this report we describe our submission to Competition 3 of Biomag2016. The proposed method is based on the work described in [1, 2]. The results reported in Table 1 show that decoding of happy faces from MEG data can be very accurate. Since the time window of each

[7]ROC AUC is an effective measure here because it is insensitive to imbalanced binary problems.

trial that resulted in highest ROC AUC was $[120ms, 1320ms]$ from stimulus onset, we expect that task-related information came from visual processing and premotor/motor activation.

expectation on the train set and bias We expect to obtain slightly inferior results on the test set, from our submission to the Competition, because some parameters of our analysis were decided without nested cross-validation on the train set (see [5] for a more detailed explanation of this issue). Moreover, the estimation of ROC AUC on the train set did not take into account the block structure of the stimulation protocol. Nevertheless we expect the impact of optimistic bias to be modest.

Possible future improvements As future work, we foresee a number of extensions, like: analyzing all subjects jointly, instead of individually, experimenting more classification algorithms, in order to investigate the structure of the SCM distribution on the tangent space, comparing with other decoding methods to study the efficacy of the proposed one. For this last aspect, the Competition will provide an interesting assessment.

Acknowledgment

We thank Alexandre Barachant for the useful suggestions and corrections to this document.

References

[1] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten. Multiclass Brain-Computer Interface Classification by Riemannian Geometry. *IEEE Transactions on Biomedical Engineering*, 59(4):920–928, April 2012.

[2] Alexandre Barachant, Stéphane Bonnet, Marco Congedo, and Christian Jutten. Classification of covariance matrices using a Riemannian-based kernel for BCI applications. *Neurocomputing*, 112:172–178, July 2013.

[3] Marcel Berger. *A Panoramic View of Riemannian Geometry*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2003.

[4] Yilun Chen, Ami Wiesel, Yonina C. Eldar, and Alfred O. Hero. Shrinkage Algorithms for MMSE Covariance Estimation. *IEEE Transactions on Signal Processing*, 58(10):5016–5029, October 2010.

[5] E. Olivetti, A. Mognon, S. Greiner, and P. Avesani. Brain decoding: Biases in error estimation. In *Brain Decoding: Pattern Recognition Challenges in Neuroimaging (WBD), 2010 First Workshop on*, volume 0, pages 40–43, Los Alamitos, CA, USA, August 2010. IEEE.